

TP1 0

Hebbar Anness

ACP Menages

(1) Loading Data

```
Men <- read.table("menages.txt", header = TRUE)
summary(Men)

##      Menage      Pain      Legume      Fruit      Viande
## Length:12      Min.   -1293.0      Min.   : 428.0      Min.   :1341.0
## Class:character  1st Qu.:381.8      1st Qu.: 596.8      1st Qu.:382.8
## Mode:character  Median:1422.0      Median: 733.0      Median:1453.5
##                Mean   :1446.7      Mean   : 732.0      Mean   :1505.0
##                3rd Qu.:1519.8      3rd Qu.: 802.5      3rd Qu.:1576.8
##                Max.   :1655.0      Max.   :1097.0      Max.   :1887.0
##
##      Viande      Volaille      Lait
## Min.:1437.7      Min.   : 526.0      Min.   :1235.0      Min.   :1258.0
## 1st Qu.:1320.2      1st Qu.: 564.8      1st Qu.:1246.0      1st Qu.:1310.2
## Median:1385.2      Median: 760.5      Median:1321.5      Median:1385.0
## Mean :1389.7      Mean   : 853.2      Mean   :1358.2      Mean   :1368.6
## 3rd Qu.:1428.5      3rd Qu.: 982.2      3rd Qu.:1434.2      3rd Qu.:1418.8
## Max. :12630.0      Max.  :1167.0      Max.  :1561.0      Max.  :1486.0

head(Men)

##      Menage Pain Legume Fruit Viande Volaille Lait Vin
## 1 BM2 332 428 354 1437 526 247 427
## 2 BM2 293 559 388 1527 567 239 258
## 3 CM2 372 767 562 1948 927 235 433
## 4 H03 406 563 341 1507 544 324 407
## 5 BM3 283 608 1503 1501 558 319 363
## 6 CA3 438 843 689 2345 1148 243 341

attach(Men)

Separation des labels en CSP en NE

CSP <- as.factor(substr(Menage,1,2)) #enleve dans chaine chaque les 2 premiers caracteres
NE <- as.factor(substr(Menage,3,3)) #enleve dans chaine chaque le 3eme caractere

Remarque -> row.names contient 1,2,... = R default

#Menages -> labels individuels
row.names(Men) <- Menage #change les noms pour ceux de menage
Men2 <- data.frame(Men, row.names=1)
head(Men2)

##      CSP NE Pain Legume Fruit Viande Volaille Lait Vin
## H02 H0 2 332 428 354 1437 526 247 427
## BM2 BM 2 293 559 388 1527 567 239 258
## CA2 CA 2 372 767 562 1948 927 235 433
## H03 H0 3 406 563 341 1507 544 324 407
## BM3 BM 3 286 608 1503 1501 558 319 363
## CA3 CA 3 438 843 689 2345 1148 243 341

summary(Men2) #affiche quantitatifs/qualitatifs = factor

##      CSP NE Pain Legume Fruit Viande
## CA14 213      Min.   -1293.0      Min.   : 428.0      Min.   :1341.0      Min.   :1437
## BM14 313      1st Qu.:381.8      1st Qu.: 596.8      1st Qu.:382.8      1st Qu.:1322
## H014 413      Median:1422.0      Median: 733.0      Median:1453.5      Median:1852
## 513      Mean   :1446.7      Mean   : 732.0      Mean   :1505.0      Mean   :1887
## 3rd Qu.:1519.8      3rd Qu.: 802.5      3rd Qu.:1576.8      3rd Qu.:2128
## Max.   :1655.0      Max.   :1097.0      Max.   :1887.0      Max.   :12630
##
##      Volaille      Lait      Vin
## Min.   : 526.0      Min.   :1235.0      Min.   :1258.0
## 1st Qu.: 564.8      1st Qu.:1246.0      1st Qu.:1310.2
## Median : 760.5      Median:1321.5      Median:1385.0
## Mean : 853.2      Mean   :1358.2      Mean   :1368.6
## 3rd Qu.: 982.2      3rd Qu.:1434.2      3rd Qu.:1418.8
## Max. :11667.0      Max.  :1561.0      Max.  :1486.0

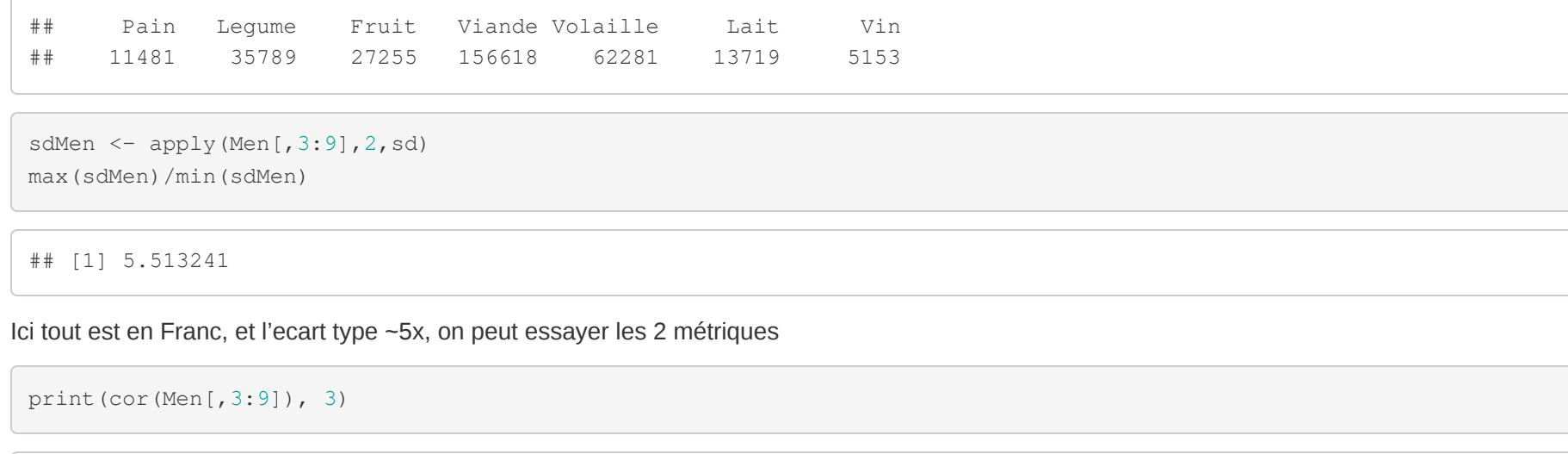
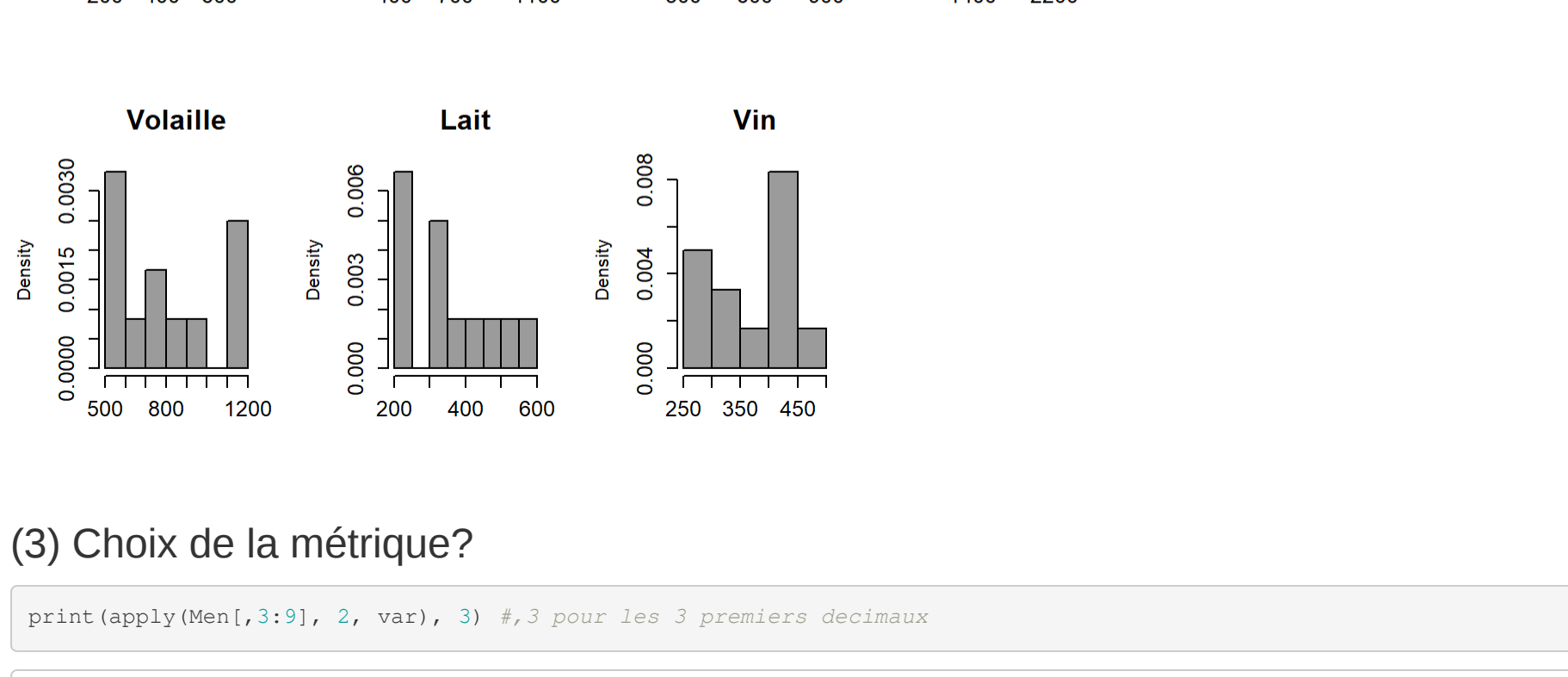
Men <- Men2
save(Men, file = "Menages.Rdata")
rm(Men2)
```

(2) ETUDE PRELIMINAIRE

```
summary(Men)

##      CSP NE Pain Legume Fruit Viande
## CA14 213      Min.   -1293.0      Min.   : 428.0      Min.   :1341.0      Min.   :1437
## BM14 313      1st Qu.:381.8      1st Qu.: 596.8      1st Qu.:382.8      1st Qu.:1322
## H014 413      Median:1422.0      Median: 733.0      Median:1453.5      Median:1852
## 513      Mean   :1446.7      Mean   : 732.0      Mean   :1505.0      Mean   :1887
## 3rd Qu.:1519.8      3rd Qu.: 802.5      3rd Qu.:1576.8      3rd Qu.:2128
## Max.   :1655.0      Max.   :1097.0      Max.   :1887.0      Max.   :12630
##
##      Volaille      Lait      Vin
## Min.   : 526.0      Min.   :1235.0      Min.   :1258.0
## 1st Qu.: 564.8      1st Qu.:1246.0      1st Qu.:1310.2
## Median : 760.5      Median:1321.5      Median:1385.0
## Mean : 853.2      Mean   :1358.2      Mean   :1368.6
## 3rd Qu.: 982.2      3rd Qu.:1434.2      3rd Qu.:1418.8
## Max. :11667.0      Max.  :1561.0      Max.  :1486.0

par(mfrow=c(1,4))
for (i in 3:9){
  hist(Men[,i], col = "#4682B4",
       main=i, ylab="freq",
       freq = FALSE,
       main=colnames(Men)[i])
}
```



(3) Choix de la métrique?

```
print(apply(Men[,3:9], 2, vapply, 3) #3 pour les 3 premiers decimaux

##      Pain Legume Fruit Viande Volaille Lait Vin
## 1 BM1 3789 2725 15618 6281 13719 5153

ndMen <- apply(Men[,3:9], 2, nd)
max(isdMen)/min(isdMen)

## [1] 5.513241
```

Idi tout est en Franc, et l'ecart type -> on peut essayer les 2 métriques

```
print(cor(Men[,3:9]), 3)

##      Pain Legume Fruit Viande Volaille Lait Vin
## Pain 1.000 0.593 0.593 0.593 0.593 0.593 0.593
## Legume 0.593 1.000 0.856 0.881 0.827 0.66280 -0.38467
## Fruit 0.196 0.856 1.000 0.959 0.926 0.33219 -0.48428
## Viande 0.821 0.881 0.959 1.000 0.982 0.37459 -0.43724
## Volaille 0.248 0.827 0.926 0.982 1.000 0.23289 -0.40016
## Lait 0.856 0.662 0.332 0.375 0.233 1.00000 0.00688
## Vin 0.384 -0.384 -0.484 -0.437 -0.400 0.00688 1.00000
```

Remarque : cor maximale (Viande, Volaille) = 98% => INTERPRETER

(4) ACP avec package ade4

```
library(ade4)

# warnings: le package "ade4" a été compilé avec la version R 4.4.3

#dudi.pca

## déchargement du serveur d'aide httpd ... fini

Remarque : check value = objet retourné + scale = TRUE ACP normée (default) + scale = FALSE ACP non normée = distance Euclidienne usuelle

On commence par l'ACP avec la distance usuelle : on garde l'unité (Franc)

acc <- dudi.pca(Men[,3:9], scale = FALSE, scanf = F, nf=7) #ACP non normée
class(acc)

## [1] "pca" "dudi"

names(acc)

## [1] "cab" "cwt" "lwt" "oig" "rank" "m" "cl" "il" "co" "il"
## [11] "call" "cent" "knn"

head(acc$li) # principal component n° 3 premières axes

##      Axis1 Axis2 Axis3 Axis4 Axis5 Axis6 Axis7
## H02 635.0477 -120.886878 -21.140067 -68.97471 9.495757 2.429738 -7.157761
## BM2 480.5551 -142.272319 12.379782 34.94484 -11.539892 12.571473 6.267031
## CA2 -132.0290 -139.746288 -61.860078 44.18650 100.396618 -23.110962 -1.789687
## H03 520.0103 12.054023 2.847675 -13.70342 -3.522482 -9.615758 16.742346
## BM3 485.9355 1.171034 65.700818 11.508814 27.183789 4.957304 -0.076140
## CA3 -588.1654 -188.443063 -71.849232 28.55981 12.143065 40.117954 3.792687

par(mfrow=c(1,2))

# Cercle principal (adapter libellé taille des noms de var)
s.cercle(coc, label="0.8", sub="Cercle de corrélation")
title("ACP NON NORMEE : cercle 1-4")
#next ... si besoin
s.cercle(coc, yax=0.8, label="0.8")
title("ACP NON NORMEE : cercle 1-3")
```

(5) Valeurs Propres et inerties cumulées

```
inertia.dudi
inertia<- inertia.dudi(acc, col.inertia = F, row.inertia = T)
names(inertia)

## [1] "tot.inertia" "row.contrib" "row.abs" "row.rel" "row.cum"
## [6] "nd" "col"

inertia$tot.inertia

##      inertia cum cum%
## Ax1 251928.43628 251928.4 88.00354
## Ax2 2424.63252 276143.1 96.46219
## Ax3 2712.39011 281676.0 96.46801
## Ax4 2108.24580 283984.3 99.20127
## Ax5 1516.02062 285900.3 99.87057
## Ax6 310.13385 286210.5 99.97883
## Ax7 60.32699 286270.8 100.00000
```

Remarque : Val propres et % inerie cumulées (ratio) => cf cours, sortie standard d'une ACP

etouts des valeurs propres

```
scatterplot.eigen(acc$eig, nf=3, box=T, sub = " ") #from ade4
```

